Financial Econometrics | Lecture 3

Linear Regression with One Regressor: Reading and Stress-Testing a Slope

In Class Group Activity

· SJTU, Summer School 2026 · Fatih Kansoy

Lesson of the day. A regression slope is only as trustworthy as the data behind it. Read it with units, separate fit from significance from size, plot before you trust, and ask whether $\mathbb{E}(u \mid X) = 0$ before you call it causal.

Phase 1 – Concept checks

ANSWERS

ConcepTest 1. **Answer.** Option **(c)**. With a 0/1 dummy, the fitted value at $D = 0$ is the intercept (13.53, the high-school mean) and at $D = 1$ it is intercept plus slope (21.34, the college mean), so the slope 7.82 is exactly the difference in means, and the test is the same difference-in-means t from Lecture 2. Option (a) invents special machinery; (b) mislabels the intercept as the premium; (d) is simply wrong.

ConcepTest 2. **Answer.** Option **(c)**. Fit, significance and size are three different things. Age explains only 1.4% of the wage spread (low R^2), yet with 13,201 observations the slope is estimated sharply and its t is about 14. Options (a) and (b) tie a low R^2 to unimportance or insignificance; (d) doubts the arithmetic instead of the intuition.

A small R^2 with a large t is the normal state of a finance regression: one factor explains little, but the relationship is still real.

ConcepTest 3. **Answer.** Option **(b)**. The slope is precise (nearly 18,000 workers) and significant, but "significant" is not "causal". Height proxies for omitted factors such as adolescent health, strength and occupation, which sit in u and move with height, so $\mathbb{E}(u \mid X) \neq 0$. Option (a) reads significance as causation; (c) confuses a low R^2 with no relationship; (d) wrongly equates zero correlation with $\mathbb{E}(u \mid X) = 0$.

Phase 2 – The desk problem

WORKED SOLUTIONS

(a) $\overline{X} = \dfrac{0+2+4+6+8+10}{6} = 5, \quad \overline{Y} = \dfrac{3+1+8+11+10+18}{6} = \dfrac{51}{6} = 8.5$. The completed table:

Month	$X - \overline{X}$	$Y - \overline{Y}$	$(X - \overline{X})(Y - \overline{Y})$	$(X - \overline{X})^2$
1	-5	-5.5	27.5	25
2	-3	-7.5	22.5	9
3	-1	-0.5	0.5	1
4	1	2.5	2.5	1
5	3	1.5	4.5	9
6	5	9.5	47.5	25
Sum	0	0	105	70

Beta: $\widehat{\beta}_1 = \dfrac{105}{70} = \mathbf{1.5}$. Alpha: $\widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1\overline{X} = 8.5 - 1.5(5) = \mathbf{1.0}$. So $\widehat{Y} = 1.0 + 1.5X$. Predict at $X = 12$: $\widehat{Y} = 1.0 + 1.5(12) = \mathbf{19}\%$. Residual for Month 1: $\widehat{u}_1 = Y_1 - \widehat{Y}_1 = 3 - (1.0 + 1.5 \cdot 0) = 3 - 1 = +\mathbf{2}$.

(b)

(1) The beta is $1.5 > 1$, so the stock is *more* volatile than the market: it tends to amplify a market move by about one and a half times. A 1% market move is matched by about a 1.5% stock move, on average.

(2) Dividing every X by 100 multiplies the slope by 100, so the beta would be reported as 150 (per unit of decimal return). The R^2 is *unchanged*: a change of units rescales the slope but not the fit.

(3) The alpha is the stock's expected excess return when the market's excess return is zero, the part of the return not explained by market exposure. In finance this is Jensen's alpha, and a positive alpha is what an active manager claims to deliver.

(c) Stata: `regress Y X, robust`. Python: `smf.ols('Y ~ X', d).fit(cov_type='HC1')`.

Phase 3 – The case

REVEAL AND ANSWERS

---

1. A worker one inch taller earns, on average, about \$708 more per year. Per centimetre: $708/2.54 \approx \$\mathbf{279}$ per cm.

2. No. A low R^2 means height explains little of the wide variation in pay, but that is about fit, not truth. With nearly 18,000 workers the slope is estimated very precisely, so it is both real and highly significant. Fit and significance are different questions.

3. Any two plausible omitted factors that move with height and affect pay: adolescent health and nutrition; physical strength (which matters in some jobs); occupation and industry (taller people are over-represented in some roles); confidence or social treatment; region and upbringing. Each sits in u and is correlated with height, so $\mathbb{E}(u \mid X) \neq 0$.

4. Instructor reveal. The slope is real and precise, but it is *not* the causal effect of height. Height proxies for adolescent health, strength and the occupations people end up in. The subgroup gap (about \$511 per inch for women, \$1,307 for men) shows the height-pay association is much stronger among men, but it does *not* prove that height pays men more: the gap is itself confounded by the different jobs men and women hold and by what height proxies for in each group. The honest reading is that a precisely estimated, significant slope can still be entirely non-causal. The fix is to control for the omitted factors, which is multiple regression, next on the syllabus.

Reserve alternate (Malta). The same lesson via a single influential point: the 65-country growth-on-trade slope is 2.31 ($t = 3.5$) with all countries, but drops to 1.68 ($t = 1.9$, no longer significant at 5%) once Malta, a re-export island far out in X, is removed. A "significant" result that rests on one observation is fragile; plot first and report with and without.

End of worked solutions.